Decision-Making Support for Adaptive Learning Management Systems based on Bayesian Inference

Nelly Bencomo
AIHS, CS, Durham university
Durham, UK
nelly@acm.org

Huma Samin
Aston University
Birmingham, UK
h.samin@aston.ac.uk

Jaime Pavlich-Mariscal
Pontificia Universidad
Javeriana
Bogotá, Colombia
jpavlich@averiana.edu.co,

ABSTRACT

A novel approach will be applied to the domain of virtual education, which involves an adaptive learning management system using Bayesian Learning. The student's progress is considered partially observable based on what has been monitored. The acquired skills by students are monitored by taking into account the results obtained from each activity performed by the student. Bayesian learning and Partially Observable Decision Processes (POMDPs) are used to guide and adapt (with the use of interventions) the learning plans according to the needs and individual characteristics of the students and their learning progress.

Keywords

Learning management system, Reinforcement Learning, POMDP, Bayesian inference, Decision making, Uncertainty

1. INTRODUCTION

Learning management systems (LMS)[19] are becoming ubiquitous due to the fact that both feedback and assessment are increasingly autonomously performed by software-based systems. Further, distance education is increasingly common in educational systems around the world [6], which has been more evident and accentuated by the COVID global crisis during 2020-21 [3, 20].

LMS usually use context to provide the students with autonomously adapted learning plans to provide a good learning experience based on both the progress made by the student and their own abilities and traits. The decision-making of the LMS can demarcate the student's characteristics and context attributes (e.g. learning style and personality). Based on the latter, LMS can provide students with a tailored set of learning activities.

In this paper, we argue that the decision-making process in LMS can be improved by using autonomous re-appraisal and updating the priorities associated with the skills to learn while using Bayesian inference and Reinforcement Learning techniques, such as Partially Observable Markov Decision Process (POMDPs)**??**.

The work presented in this paper is motivated by a case study in the area of virtual education. In the case study, an extension of the Multi-Reward Partially Observable Markov Decision Process (MR-POMDP) [14, 15, 11] is used to autonomously adapt the information that is provided to the student when their performance in a particular skill is either not at the desired level or is higher than expected. Due to changes in the context and progress of the learning process, the relevance of the information provided to the student, the learning activities and skills to be acquired may be affected. The information and plan are relevant only if they help the student to achieve the learning goals (e.g. developing a new skill). Changes in the relevance of information and activities mean changes in the priorities associated. An advantage is that the learning plan and the priorities implied are adapted according to the current characteristics, progress and needs of the student and his/her learning progress. In the way how we approach the domain problem, signalling how relevant some learning objectives are is done by using weights and or preferences used by the multi-objective decision-making process [11] to decide on the adapted learning plans, which correspond to adaptations.

2. MOTIVATING SCENARIO

The main goal of the LMS considered in Fig 1 based on [7] is to use context information to perform dynamic adaptive planning that best suits the student's requirements according to their own performance, progress and given traits.

The process starts with the creation of an adaptive plan. At this initial stage, the instructor defines a set of skills that need to be acquired by the students by the end of a virtual course. To achieve each skill, a set of learning plans are created that are composed of several academic activities.

When the student interacts with the LMS for the first time, the application will use context information such as the student's learning styles and personality traits. As is depicted in Fig. 1, this information is measured by using a set of predefined tests. With the results of the tests, the application creates a user profile. After, the profile is used to select the most suitable plan for the student.

A relevant issue in the case study is that the pertinence

of the learning plan, based on a specific context, is usually determined before the study starts. This task is usually performed by experts in education and pedagogy based on general knowledge that may not completely agree with the traits of individual students [4]. In a traditional non-autonomous setting, when the experts detect that the learning plan is not helping to improve the required skill anymore (e.g. when the student gets a lower grade), the plan needs to be adjusted to ensure that the information provided to the student is still relevant for helping to acquire the skill. Adjusting and monitoring the execution of a learning plan for a single student may be a complex and time-consuming activity. Therefore, the process of adapting and monitoring, for instance, 20 different plans for a group of 20 heterogeneous students may turn out to be even more challenging. We argue that using the approach based on [11] if the expected utility value is not the targeted value, the application can re-adjust autonomously the tailored learning plan to be provided to the student to improve the learning experience. This process is scalable, as it can be done for both a single student and a large group of them.

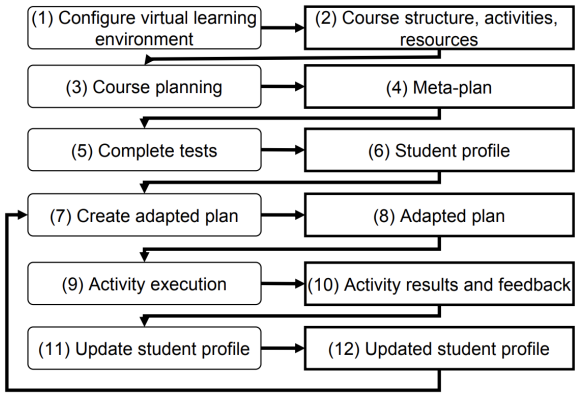

Figure 1: LMS Process (from [7])

3. PARTIALLY OBSERVABLE MARKOV DECISION PROCESSES (POMDPS)

This section describes the baseline concepts of single-objective and multi-objective POMDPs. POMDPs [16, 17] are Reinforcement Learning techniques used to solve sequential decision-making problems under uncertainty. In order to incorporate uncertainty about the state of the environment, POMDPs consider the decision-making agents working in a partially observable environment.

Single and Multi-objective POMDPS

A single-objective POMDP is specified as a tuple:

$$<\text{S, A, Z, T, O, R, } \gamma >$$

where S represents the set of states referring to a description of the state of the environment; A represents the set of Actions that the decision-making agent can choose to perform at a particular point of time; Z represents the set of Observations describing the information received by the decision-making agent using sensors associated with the set of states S; T is the transition function $T(s, a, s') = P(s'|s, a)$ representing the probability of moving to the next state s' given an action a and current state s; O represents the observation function $O(s, a, z) = P(z|s, a)$ referring to the probability of observing the observation z given an action a and resultant state s; R is the reward function R(s,a) specifying a scalar real value generated by the environment as a result of the action a taken by the decision-making agent given the state s of the environment; γ is the discount factor. The diagramatic representation of POMDP as presented in [8] is shown in Fig 2. A policy π, a mapping from the state of the environment to action, is found by the decision-making agent such that it maximizes the value function i.e. the expected utility value of the sum of discounted rewards as follows:

$$V_\pi = E_\pi[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2}...|s_t] \quad (1)$$

Due to the partially observable nature of states in a POMDP, a belief b over the *hidden state* of the environment is maintained. Hence, the value function V_b is defined in terms of the belief and can be represented by a set of α vectors A [16, 12]. Each α vector, associated with an action a, has a length of $|S|$ providing a value for each state s is computed as follows:

$$\alpha_a = [V(s_i), V(s_{i+1}), ..., V(s_{(n)})] \quad (2)$$

Here $V(s_i)$ represents the value of the value function for state s_i provided the n number of states in total.

Thus, the value of the belief given A is computed as:

$$V_b = \max_{\alpha \in A} b.\alpha \quad (3)$$

Therefore, for each belief b, a set of α vectors A provides a policy π_A for the selection of the action that maximizes the value.

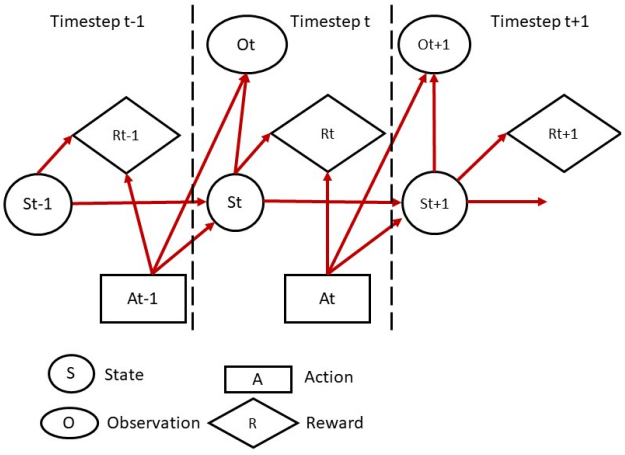

Figure 2: POMDP process

In comparison to single-objective POMDPs, multi - objective POMDPs [10] are POMDPs with more that one reward

value. Instead of scalar reward, multi-objective POMDPs have a vector-valued reward function. The size of the reward vector is equal to the number of objectives. Each single element in the reward vector is associated with each individual objective. As reward is a vector in multi-objective POMDPs, the value function given an initial belief is also a vector. Hence, the expected utility value for each objective is separately computed during the decision-making process in case of multi-objective POMDPs.

4. POMDPS FOR DECISION-MAKING SUPPORT FOR LMS

It is possible to cast the decision-making of an LMS as a POMDP. The main contribution of this paper is framing the decision-making for LMS as that of a POMDP. The graph-based structure of a POMDP matches that of the decision-making of an LMS.

The hidden state correspond to the level of skills met by the student at any point of the learning process, which is partially observable based on the grades obtained by the student. This is because we cannot be 100% sure but hold a belief about how well a student has reached the skills based on the marks obtained. Therefore, based on the mathematical model of the POMDP, the partially observable state of the POMDP corresponds to the belief b about how well the student skills have been met. Moreover, the extent to which the skill has been reached by the student is monitored based on the evaluations that yield grades or mark. This is dictated by the observation model. The actions in the POMDP correspond with the activity plans in the LMS. The time steps are mapped to an activity execution and evaluations in the LMS. The execution and evaluation of these activities and assessments may have different temporal frequencies (e.g. weekly or daily).

As the states associated with students' skills evolve based on the learning process, the plans currently assigned to them may not be suitable anymore. To address this situation, the POMDP would adapt the student's plan, which means that whenever students change their state, the POMDP would create and suggest a new plan to better suit the new conditions and context of the student.

When performing the above, the priorities associated with the activities to be chosen may need to be changed based on the evidence collected about the current state of skills acquired by the student (e.g. some activities that initially were considered to have a higher priority may not be critical anymore or the opposite). Accordingly, the LMS empowered by the POMDP would autonomously change the priorities. Therefore, the decision support would autonomously adjust or suggest the pertinence of alternative activities when evidence exists that the student has not been able to reach a required skill.

4.1 Scenario

A scenario we have worked out using our POMDP solver [11], is the following:

Let's assume a simple example of three learning activities (*essay, simulation*, and *presentation*) which help to acquire the skills *critical-thinking*, *assertive communication* and *problem-solving* with different preferences and transition probabilities associated.

Let's also assume that the student has interacted with the LMS and has performed a set of suggested activities. At the end of these activities, the evaluation indicates that his/her grade for the skill *problem-solving* is still at the basic level. This means that in the next iteration, the application should suggest an activity that should help strengthen the given skill. In order to do that, the initial preferences for activities given the current state may need to be re-assessed, by using multi-objective POMDPs [11] by the computation of expected utilities as described in Section 3. Multi-objective POMDPs allow the reasoning of the individual preferences related to the activities, in order to reach the level of skills targeted. The new preferences suggest that in the next iteration the student shouldn't perform the *essay* activity (suggested by the initial preferences) but the *simulation* activity, which according to what the experts have established, is an alternative activity that also helps to develop the skill *critical thinking*. Having a lecturer to do this level of individualisation would require many hours of work. However, the POMDP solver is able to offer such a level in just seconds.

4.2 Challenges

The kind of individualisation [9] focus offered by a POMDP does not come for free, as the elicitation of the transition model, the observation model, and the initial preferences require work and expertise.

As shown in Section 3, the transition model represents the probability to move to the next state *s'* given an action *a* and current state *s*. In other words, and for the case of the LMS, it is the probability to acquire a new skill *sk'* by undertaking the learning activity *la* and the current skill *sk*.

The observation function $O(s, a, z) = P(z|s, a)$ refers to the probability of observing the observation z given an action a and the resulting state s, which in the LMS corresponds with the observation of the mark m achieved given the skill acquired *sk* and the learning activity *la*.

The elicitation of these probabilities requires new techniques [18, 5].

We also would like to explore the use of Bayesian inference to study the process of learning skills to define singular profiles that describe the generic learning skills of students based on their previous performance. In other words, using Bayesian inference, the LMS system could infer (i.e. learn) and therefore, make conclusions about the performance of students in the future, including other modules. This is an exciting research line that we did not foresee before. We consider the research venue CausalEDM'22 an excellent one to get feedback to pursue further our research on Reinforcement Learning techniques such as POMDPS for Education [13]. We believe that our work on decision-making under uncertainty for autonomous, self-adaptive systems[11, 2, 1], is of great value for the Artificial Intelligence in Education (AIE) community.

5. REFERENCES

[1] Nelly Bencomo and Amel Belaggoun. Supporting decision-making for self-adaptive systems: From goal models to dynamic decision networks. In Joerg Doerr and Andreas L. Opdahl, editors, *Requirements Engineering: Foundation for Software Quality*, pages 221–236, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[2] Nelly Bencomo, Amel Belaggoun, and Valérie Issarny. Dynamic decision networks for decision-making in self-adaptive systems: a case study. In Marin Litoiu and John Mylopoulos, editors, *Proceedings of the 8th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS 2013, San Francisco, CA, USA, May 20-21, 2013*, pages 113–122. IEEE Computer Society, 2013.

[3] Alfiya R. Masalimova, Maria A. Khvatova, Lyudmila S. Chikileva, Elena P. Zvyagintseva, Valentina V. Stepanova, and Mariya V. Melnik. Distance learning in higher education during covid-19. *Frontiers in Education*, 7, 2022.

[4] James McMillan. *lassroom Assessment: Pearson New International Edition: Principles and Practice for Effective Standards-Based Instruction*. Pearson Higher, 2013.

[5] A. O'Hagan, C. E. Buck, J. R.and Garthwaite P. H.and Jenkinson D. J. Daneshkhah, A.and Eiser, J. E. Oakley, and T. Rakow. *Uncertain Judgements: Eliciting Expert Probabilities*. Wiley, 2006.

[6] Ashish Pant. Distance learning: History, problems and solutions. *Computer Science and Information Technology (ACSIT)*, 1(2):65–70, 2014.

[7] Jaime Pavlich-Mariscal., Yolima Uribe., Luisa Fernanda Barrera León., Nadia Alejandra Mejia-Molina., Angela Carrillo-Ramos., Alexandra Pomares Quimbaya., Monica Brijaldo., Martha Sabogal., Rosa Maria Vicari., and Hervé Martin. Ashyi-edu: Applying dynamic adaptive planning in a virtual learning environment. In *Proceedings of the 7th International Conference on Computer Supported Education - CSEDU,*, pages 52–63. INSTICC, SciTePress, 2015.

[8] David Poole and Alan Mackworth. *Artificial Intelligence - foundations of computational agents - Partially Observable Decision Processes*. Cambridge University Press, 2017.

[9] Ethan Prihar, Aaron Haim, Adam Sales, and Neil Heffernan. Automatic interpretable personalized learning. In *Proceedings of the Ninth ACM Conference on Learning @ Scale*, L@S '22, page 1–11, New York, NY, USA, 2022. Association for Computing Machinery.

[10] Diederik Marijn Roijers, Shimon Whiteson, and Frans A. Oliehoek. Point-Based Planning for Multi-Objective POMDPs. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, June 2015.

[11] Huma Samin, Nelly Bencomo, and Pete Sawyer. Decision-making under uncertainty: be aware of your priorities. *Software and Systems Modeling (SoSyM)*, 2022.

[12] Guy Shani, Joelle Pineau, and Robert Kaplow. A survey of point-based pomdp solvers. *Autonomous Agents and Multi-Agent Systems*, 27(1):1–51, 2013.

[13] Adish Singla, Anna N. Rafferty, Goran Radanovic, and Neil T. Heffernan. Reinforcement learning for education: Opportunities and challenges, 2021.

[14] Harold Soh and Yiannis Demiris. Evolving policies for multi-reward partially observable Markov decision processes (MR-POMDPs). In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 713–720, January 2011.

[15] Harold Soh, Yiannis Demiris, Harold Soh, and Yiannis Demiris. Multi-Reward Policies for Medical Applications: Anthrax Attacks and Smart Wheelchairs. In *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*, pages 471–478, 2011.

[16] M. T. J. Spaan and N. Vlassis. Perseus: Randomized Point-based Value Iteration for POMDPs. *Journal of Artificial Intelligence Research*, 24:195–220, August 2005. arXiv: 1109.2145.

[17] Matthijs TJ Spaan. Partially observable markov decision processes. In *Reinforcement Learning*, pages 387–414. Springer, 2012.

[18] Alistair Sutcliffe, Pete Sawyer, Gemma Stringer, Samuel Couth, Laura JE Brown, Ann Gledson, Christopher Bull, Paul Rayson, John Keane, Xiao-jun Zeng, et al. Known and unknown requirements in healthcare. *Requirements engineering*, 25(1):1–20, 2020.

[19] Darren Turnbull, Ritesh Chugh, and Jo Luck. Learning management systems: a review of the research methodology literature in australia and china. *International Journal of Research & Method in Education*, 44(2):164–178, 2021.

[20] Ewelina Zarzycka, Joanna Krasodomska, Anna Mazurczak-Mąka, and Monika Turek-Radwan. Distance learning during the covid-19 pandemic: students' communication and collaboration and the role of social media. *Cogent Arts & Humanities*, 8(1):1953228, 2021.